

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-175530

(43)公開日 平成11年(1999) 7月2日

(51)Int.Cl.⁶

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/401

3 1 0 D

審査請求 未請求 請求項の数11 O L (全 8 頁)

(21)出願番号 特願平9-337069

(22)出願日 平成9年(1997)12月8日

(71)出願人 000004226

日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号

(72)発明者 杉崎 正之

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者 大久保 雅且

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者 田中 一男

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74)代理人 弁理士 若林 忠 (外2名)

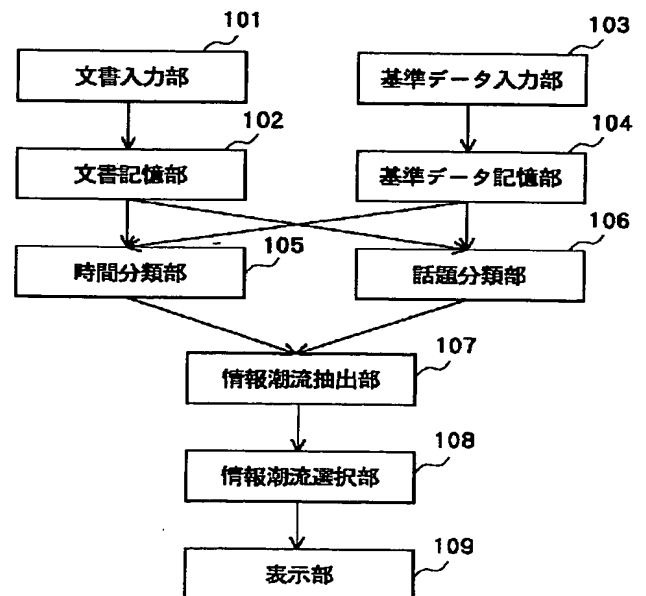
最終頁に続く

(54)【発明の名称】 情報潮流提示方法および装置ならび情報潮流提示プログラムを記録した記録媒体

(57)【要約】

【課題】 複数の情報発信媒体から発信される情報から利用者が自分に必要な情報を容易に得られるようにする。

【解決手段】 複数の情報発信媒体から発信された情報と分類のための基準データをそれぞれ文書入力部101、基準データ入力部103より入力し、それぞれ文書記憶部102、基準データ記憶部104に記憶する。時間分類部105では入力された文書を時間に関する分類カテゴリで分類する。話題分類部106では話題に関する分類カテゴリで、入力された文書を分類する。情報潮流抽出部107では、時間に関する分類カテゴリに分類された各文書に含まれる、話題に関する各カテゴリを時系列に示す情報潮流を情報発信媒体毎に抽出する。情報潮流選択部108では抽出された情報潮流を比較し、利用者が希望する情報潮流を選択し、表示部109により使用者に提示する。



【特許請求の範囲】

【請求項 1】 複数の情報発信媒体から発信された、発信された時間の情報を有する大量の文書から利用者が希望する特定の情報の潮流を抽出し、利用者に提示する情報潮流提示方法であって、
 複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力し、記憶装置に記憶する段階と、
 前記文書を分類する基準となる分類基準データと利用者が希望する情報の潮流を選択するための選択基準データを入力し、記憶装置に記憶する段階と、
 前記分類基準データを用いて時間に関する 1 以上の分類カテゴリを作成し、該分類カテゴリで前記文書を前記情報発信媒体毎に分類する段階と、
 前記分類基準データを用いて話題に関する 1 以上の分類カテゴリを作成し、該分類カテゴリで前記文書を前記情報発信媒体毎に分類する段階と、
 時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分類された文書の数求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する段階と、
 前記各情報発信媒体の情報潮流を比較し、これら情報潮流の中から前記選択基準データに基づいて特定の情報潮流を抽出する段階と、
 前記特定の情報潮流を表示装置に表示する段階を有する情報潮流提示方法。

【請求項 2】 前記時間に関する分類基準データが、分類を行う時間間隔である、請求項 1 記載の方法。

【請求項 3】 前記話題に関するカテゴリを文書中から自動的に生成する段階を有する、請求項 1 記載の方法。

【請求項 4】 前記各情報発信媒体に共通の情報潮流を特定の情報潮流として抽出する、請求項 1 記載の方法。

【請求項 5】 前記各情報発信媒体に固有の情報潮流を特定の情報潮流として抽出する、請求項 1 記載の方法。

【請求項 6】 複数の情報発信媒体から発信された、発信された時間の情報を有する大量の文書から、利用者が希望する特定の情報の潮流を抽出し、利用者に提示する情報潮流提示装置であって、
 複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力する文書入力手段と、
 前記文書入力手段で入力された文書を記憶する文書記憶手段と、
 前記文書を分類する基準となる分類基準データと利用者が希望する情報の潮流を選択するための選択基準データを入力する基準データ入力手段と、
 前記基準データ入力手段で入力されたデータを記憶する基準データ記憶手段と、
 前記分類基準データを用いて時間に関する 1 以上の分類

カテゴリを作成し、該分類カテゴリで前記文書を前記情報発信媒体毎に分類する時間分類手段と、
 前記分類基準データを用いて話題に関する 1 以上の分類カテゴリを作成し、該分類カテゴリで前記文書を前記情報発信媒体毎に分類する話題分類手段と、
 時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分類された文書の数求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する情報潮流抽出手段と、
 前記各情報発信媒体の情報潮流を比較し、これら情報潮流の中から前記選択基準データに基づいて特定の情報潮流を抽出する情報潮流選択手段と、
 前記特定の情報潮流を表示する表示手段を有する情報潮流提示装置。

【請求項 7】 前記時間に関する分類基準データが、分類を行う時間間隔である、請求項 6 記載の装置。

【請求項 8】 前記話題に関するカテゴリが文書中から自動的に生成される、請求項 6 記載の装置。

【請求項 9】 前記情報潮流選択手段は前記各情報発信媒体に共通の情報潮流を抽出する、請求項 6 記載の装置。

【請求項 10】 前記情報潮流選択手段は前記各情報発信媒体に固有の情報潮流を抽出する、請求項 6 記載の装置。

【請求項 11】 複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力し、記憶装置に記憶する処理と、
 前記文書を分類する基準となる分類基準データと、利用者が希望する情報の潮流を選択するための選択基準データを入力し、記憶装置に記憶する処理と、
 前記分類基準データを用いて時間に関する 1 以上の分類カテゴリを作成し、該分類カテゴリで前記文書データを前記情報発信媒体毎に分類する処理と、
 前記分類基準データを用いて話題に関する 1 以上の分類カテゴリを作成し、該分類カテゴリで前記文書を前記情報発信媒体毎に分類する処理と、
 時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分類された文書の数求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する処理と、
 前記各情報発信媒体の情報潮流を比較し、前記情報潮流の中から前記選択基準データに基づいて特定の情報潮流を抽出する処理と、
 前記特定の情報潮流を表示装置に表示する処理をコンピュータに実行させる情報潮流提示プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数の情報発信媒体から発信された、発信された時間の情報を有する大量の文書データから、利用者が希望する特定の情報の潮流を抽出し、利用者に提示する情報潮流提示方法および装置に関する。

【0002】

【従来の技術】近年、インターネットなどのコンピュータネットワークを通じて、電子化された大量の文書をやり取りできるようになっている。そのため、個人が必要とする情報を検索できるようなサービスがネットワーク上で実現されている。しかし、そのために自分が獲得した情報が大量になってしまい、個々の情報の持つ特徴を抽出することが困難となる。そこで、獲得した情報を分類し整理する技術が必要となる。

【0003】従来から、文書情報の自動的に分類する手法の研究が行われている。代表的な手法としては、図書館のように分類するための区切り（カテゴリと呼ぶ）が既知で、新規の情報に対してそれぞれ適切と思われるカテゴリに分類する手法（“分類体系相互の関係を利用したテキストの自動分類” 山本、増山（豊橋技術大学）内藤（NTT）、1995）や、分類するカテゴリが未知で、文書集合の中から類似する文書を集めて分類カテゴリを作成し割り当てるという方法（“競合学習ニューラルネットワークによる自動分割” 菊池、松岡ら、（宇都宮大他）、1995）、などがある。これらの技術により、大量の文書の分類処理を行っている。

【0004】また、新聞記事のような、時間情報を持つ文書を対象に、話題の遷移を抽出する手法の研究も行われている（“自動分類を用いたトレンド・アウェアネスの支援” 杉崎、井上、大久保、田中（NTT）1997）。この技術は、ある文書集合から話題が類似した文書を自動収集してカテゴリを作成し、それをある時間間隔で分割し提示することにより、ある話題に関する記事数や話題の分岐などの時間変化を抽出するものである。

【0005】

【発明が解決しようとする課題】上述した従来の方法では、新規に情報を発信している媒体（ここでは情報発信媒体と呼ぶ）から抽出した話題の時間的な変化を見て、人間がどの話題が最近流行しているかを判断しなければならない。

【0006】本発明の目的は、利用者が、複数の情報発信媒体から発信される情報の中から自分に必要な情報を容易に得ることができる情報潮流提示方法および装置ならびに情報潮流提示プログラムを記録した記録媒体を提供することにある。

【0007】

【課題を解決するための手段】本発明は情報潮流提示方法は、複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力し、記憶装置に記憶する段階と、文書を分類する基準と

なる分類基準データと利用者が希望する情報の潮流を選択するための選択基準データを入力し、記憶装置に記憶する段階と、分類基準データを用いて時間に関する1以上の分類カテゴリを作成し、該分類カテゴリで文書を情報発信媒体毎に分類する段階と、分類基準を用いて話題に関する1以上の分類カテゴリを作成し、該分類カテゴリで文書を情報発信媒体毎に分類する段階と、時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分離された文書の数を求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する段階と、各情報発信媒体の情報潮流を比較し、これら情報潮流の中から選択基準データに基づいて特定の情報潮流を抽出する段階と、特定の情報潮流量を表示装置に表示する段階を有する。

【0008】本発明の情報潮流提示装置は、複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力する文書入力手段と、文書入力手段で入力された文書を記憶する文書記憶手段と、文書を分類する基準となる分類基準データと利用者が希望する情報の潮流を選択するための選択基準データを入力する基準データ入力手段と、基準データ入力手段で入力されたデータを記憶する基準データ記憶手段と、分類基準データを用いて時間に関する1以上の分類カテゴリを作成し、該分類カテゴリで文書を情報発信媒体毎に分類する時間分類手段と、分類基準データを用いて話題に関する1以上の分類カテゴリを作成し、分類カテゴリで文書を情報発信媒体毎に分類する話題分類手段と、時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分類された文書の数を求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する情報潮流抽出手段と、各情報発信媒体の情報潮流を比較し、これら情報潮流の中から基準選択データに基づいて特定の情報潮流を抽出する情報潮流選択手段と、特定の情報潮流量を表示する表示手段を有する。

【0009】本発明の情報潮流提示プログラムを記録した記録媒体は、複数の情報発信媒体から発信された、発信された時間の情報を有する、自然言語で記述された文書を入力し、記憶装置に記憶する処理と、文書を分類する基準となる分類基準データと、利用者が希望する情報の潮流を選択するための選択基準データを入力し、記憶装置に記憶する処理と、分類基準データを用いて時間に関する1以上の分類カテゴリを作成し、該分類カテゴリで文書を情報発信媒体毎に分類する処理と、分類基準データを用いて話題に関する1以上の分類カテゴリを作成し、該分類カテゴリで文書を情報発信媒体毎に分類する処理と、時間に関する分類カテゴリで分類された各文書に含まれる、話題に関するカテゴリで分類された文書の

数を求め、時間に関する各分類カテゴリにおける話題に関する各カテゴリの文書数を時系列に示す情報潮流を各情報発信媒体毎に抽出する処理と、各情報発信媒体の情報潮流を比較し、前記情報潮流の中から選択基準データに基づいて特定の情報潮流を抽出する処理と、特定の情報潮流を表示装置に表示する処理を有する情報潮流提示プログラムを記録している。

【0010】まず、文書、例えば、記事集合から話題の遷移を抽出する。本発明で分類対象となる文書は、時間の情報を有する文書であり、その文書を作成した時間や文書内で表現している時間を利用する。記事の集合は、新聞社やラジオ局やテレビ局など、テキスト情報を発信している媒体から発信されている情報を別々に、また組み合わせ合わせて利用する。話題の時間的な変化（これを情報潮流と呼ぶ）は、記事集合を時間毎に分類した結果と話題毎に分類した結果を用いて、共通した話題のカテゴリを時間順に繋げていく。

【0011】複数の情報発信媒体より抽出された話題の時間的な変化の様子を比較することにより、ある特定の情報発信媒体だけから発信されている話題と、ほとんどの情報発信媒体が発信している話題を判断することができる。多くの情報発信媒体において、抽出された話題が同一で、かつ、話題の時間的な変化が類似しており、それらが他の情報潮流と比較して大量に記事が発信されている場合に、最も盛り上がっている話題であると判断する。

【0012】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して説明する。

【0013】図1を参照すると、本発明の一実施形態の情報潮流提示装置は、文書入力部101と文書記憶部102と基準データ入力部103と基準データ記憶部104と時間分類部105と話題分類部106と情報潮流抽出部107と情報潮流選択部108と表示部109とで構成されている。

【0014】文書入力部101では、処理を施したい文書を入力する。処理を施したい文書としては、コンピュータに入力した文書すべてが含まれる。例えば、新聞記事やインターネット上のHTMLファイルやネットニュースや文字放送やFM多重放送やテレビでの放送原稿などがある。文書記憶部102では、文書入力部101において入力した情報を、情報発信媒体それぞれ、あるいは、情報発信媒体のいくつか、あるいは、すべての情報発信媒体全体の文書を取り出せるように記憶装置に記憶する。

【0015】基準データ入力部103では、システムが処理を行う際の基準データを入力する。例えば、時間分類部105において、ある決まった時間間隔で分類を行うときの時間間隔や、話題分類部106において、分類を行うときの閾値や分類し抽出したい情報、情報潮流選

択部108において、情報潮流を比較する情報発信媒体の選択や比較する情報の指示、表示部109で表示する情報の指定などを入力する。基準データ記憶部104では基準データ入力部103において入力した基準データを記憶装置に記憶する。

【0016】時間分類部105と話題分類部106では実際の文書の分類を行う。時間分類部105では、基準データ入力部103で入力された時間間隔で分類カテゴリを作成し、文書を分類する。例えば、1997年の新聞記事に対して、基準データ入力部103で「8月から9月まで1日刻み」と入力すると、「“1997年8月1日”の分類カテゴリ」「“1997年8月2日”の分類カテゴリ」・・・「“1997年9月30日”の分類カテゴリ」を作成する。1997年8月1日の新聞の記事は「“1997年8月1日”の分類カテゴリ」に分類され、1997年10月1日の新聞の記事はどの時間分類カテゴリにも分類されない。話題分類部106では、文書入力部101において入力された文書集合内で類似する文書を収集し分類カテゴリを作成し、文書を分類する。例えば、新聞記事を文書集合としたとき「火星探査」の話題の新聞記事が多くあった場合「“火星探査”の分類カテゴリ」という新たな分類カテゴリをシステムが自動的に作成し、“火星探査”に関する新聞記事を「“火星探査”の分類カテゴリ」に割り当てる。また、あらかじめシステム利用者により与えられた分類カテゴリに対しては、該当する文書を分類する。例えば、システム利用者が“海水浴”に関する記事を集めるように指示した場合、「“海水浴”に関する分類カテゴリ」を作成し、“海水浴”に関する記事をこのカテゴリに割り当てる。

【0017】情報潮流抽出部107では、話題分類部106で求めた分類カテゴリを、時間分類部105で計算された分類結果を用いて、時間軸に沿って並べる作業を行う。例えば、1997年の新聞記事に対し、話題分類部106からは「“陸上競技”の分類カテゴリ」「“火星探査”の分類カテゴリ」を、時間分類部105からは「“1997年8月1日”の分類カテゴリ」・・・「“1997年9月30日”の分類カテゴリ」を情報潮流抽出部107が受け取る。「“1997年8月1日”の分類カテゴリ」に割り当てられた文書のうち「“陸上競技”の分類カテゴリ」「“火星探査”の分類カテゴリ」のそれぞれに分類された文書を抽出し、分類カテゴリ内の文書数を計算する。その結果を、“1997年8月1日”分から“1997年9月30日”分まで、「“陸上競技”の分類カテゴリ」「“火星探査”の分類カテゴリ」内の文書と文書数を並べる。

【0018】情報潮流選択部108では、情報潮流抽出部107で抽出された情報潮流を比較し、その差異を抽出する。例えば、2つの情報発信媒体から抽出された情報潮流を比較して、両方の情報発信媒体で共通の時期に

“火星探査”に関する情報潮流が増加していた場合、その時期には“火星探査”に関する話題が盛り上がっていることを抽出し、また、一方の情報発信媒体で“政治”に関する記事が多く発信されていて、もう一方の情報発信媒体ではあまり発信されていないければ、前者は“政治”の話題より多く発信している情報発信媒体であることを抽出する。この組み合わせは、基準データ入力部103により入力された基準データを元に計算を行う。

【0019】本実施形態の処理の流れを具体例により図2を参照して説明する。

【0020】システムに入力する文書集合は、新聞社Aと新聞社Bと新聞社Cのそれぞれの1997年の新聞記事とする。

【0021】文書集合を文書入力部101より入力し、記憶する(ステップ201、202)。次に、基準データ入力部103より、システムが分類を行うための基準データを入力し、記憶する(ステップ203、204)。例えば、あらかじめ集めたい新聞記事があれば「“交通事故”に関する記事を集める分類カテゴリ」や「“地震”に関する記事を集める分類カテゴリ」を入力する。また、時間分類部105で分類するための時間間隔を入力する。例えば、「一日刻み」や「一週間刻み」と入力する。ここでは、「一ヶ月刻み」と入力したとする。また、情報潮流選択部108で比較する情報発信媒体と比較の仕方を入力する。例えば、「新聞社Aと新聞社Bで多く発信された話題」や「各新聞社でしか見られない話題」と入力する。ここでは、「全新聞社で多く発信された話題」と入力したとする。

【0022】以上の基準データにより、時間分類部105と話題分類部106で新聞記事の分類を行う(ステップ205、206)。この結果から、時間分類部105では「“1月”の分類カテゴリ」「“2月”の分類カテゴリ」・・・と一ヶ月刻みで分類カテゴリが作成される。また、話題分類部106では、基準データ入力部103で入力された「“交通事故”に関する記事を集める分類カテゴリ」や「“地震”に関する記事を集める分類カテゴリ」などに該当する記事があれば分類を行い、それ以外の記事で、例えば、食中毒に関する記事が多数存在する場合は「“食中毒”に関する分類カテゴリ」を作成し、該当する記事を割り当てる。

【0023】次に、情報潮流抽出部107で、時間分類部105と話題分類部106の分類結果を用いて、時系列に各分類カテゴリを並べ、その結果(情報潮流)を出力する(ステップ207)。出力例を図3に示す。

【0024】次に、情報潮流選択部108で、情報潮流抽出部107の抽出結果を用いて、各情報潮流の比較を行う(ステップ208)。例えば、新聞社Aの新聞記事から抽出された情報潮流として「“食中毒”に関する情報潮流」と「“オリンピック”に関する情報潮流」と「“国会の出来事”に関する情報潮流」が、新聞社Bの

新聞記事から抽出された情報潮流として「“食中毒”に関する情報潮流」と「“オリンピック”に関する情報潮流」と「“台風”に関する情報潮流」が、新聞社Cの新聞記事から抽出された情報潮流として「“食中毒”に関する情報潮流」と「“流星”に関する情報潮流」と

「“地震”に関する情報潮流」が、情報潮流抽出部107において抽出されたとする。基準データ入力部103において基準データとして「最近、話題となっている情報を表示したい」が入力されたとすると、すべての新聞社A、B、Cから発信されている情報潮流を抽出する。この場合は、情報潮流選択部108から「“食中毒”に関する情報潮流」が抽出される。基準データ入力部103において基準データとして「各情報発信媒体で固有に話題となっている情報を表示したい」が入力されたとすると、情報潮流選択部108から新聞社Aでは「“国会の出来事”に関する情報潮流」が、新聞社Bでは「“台風”に関する情報潮流」が、新聞社Cでは「“流星”に関する情報潮流」と「“地震”に関する情報潮流」が抽出される。情報発信媒体が2件のときの情報潮流選択部108での抽出例を図4、図5に示す。

【0025】表示部109では、情報潮流抽出部107や情報潮流選択部108で得た結果を、グラフや必要な記事などを用いて表示する(ステップ209)。

【0026】図6を参照すると、本発明の他の情報潮流提示装置は入力装置301と記憶装置302と表示装置303と記録媒体304とデータ処理装置305で構成される。

【0027】入力装置301は図1中の文書入力部101と基準データ入力部103に相当する。記憶装置302は、入力された文書データと記憶する文書記憶部302aと、入力された基準データを記憶する基準データ記憶部302bを含んでいる。表示装置303はCRTなどのディスプレイで、情報潮流を表示する。記録媒体304は図2に示した処理からなる情報潮流提示プログラムを記録した、CD-ROM、FD、半導体メモリなどの記録媒体である。データ処理装置305は、記録媒体304から情報潮流提示プログラムを読み込み、実行する。

【0028】

【発明の効果】以上説明したように本発明は、大量の文書集合において、文書集合から個々の文書が持つ情報以外に、時間や話題で分類することによってのみ得ることができる情報の時間的な変化を抽出し、その結果を用いて比較計算を行うことで、システム利用者が自分に必要な情報を容易に得ることができる。

【図面の簡単な説明】

【図1】本発明の一実施形態の情報潮流提示装置の概略構成を示すブロック図である。

【図2】図1の情報潮流提示装置の処理を示す流れ図である。

【図3】図1の情報潮流提示装置の出力の一例を示す図である。

【図4】図1の情報潮流提示装置の出力の一例を示す図である。

【図5】図1の情報潮流提示装置の出力の一例を示す図である。

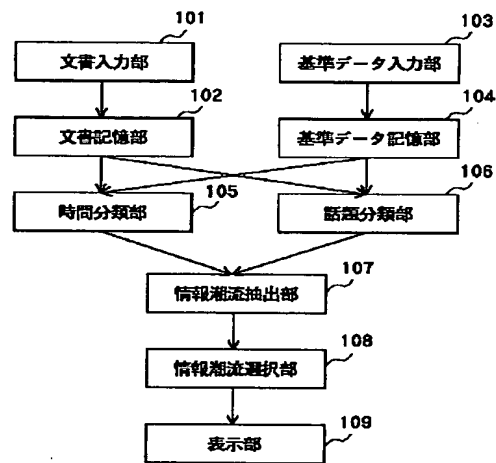
【図6】本発明の他の実施形態の情報潮流提示装置の概略構成を示すブロック図である。

【符号の説明】

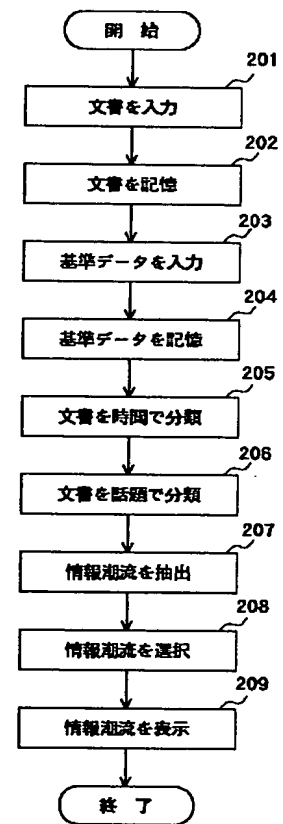
101 文書入力部
102 文書記憶部
103 基準データ入力部
104 基準データ記憶部

105 時間分類部
106 話題分類部
107 情報潮流抽出部
108 情報潮流選択部
109 表示部
201～209 ステップ
301 入力装置
302 記憶装置
302 a 文書記憶部
302 b 基準データ記憶部
303 表示装置
304 記録媒体
305 データ処理装置

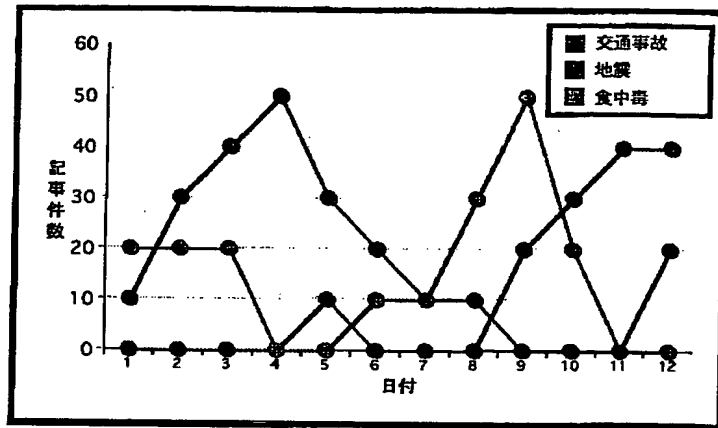
【図1】



【図2】

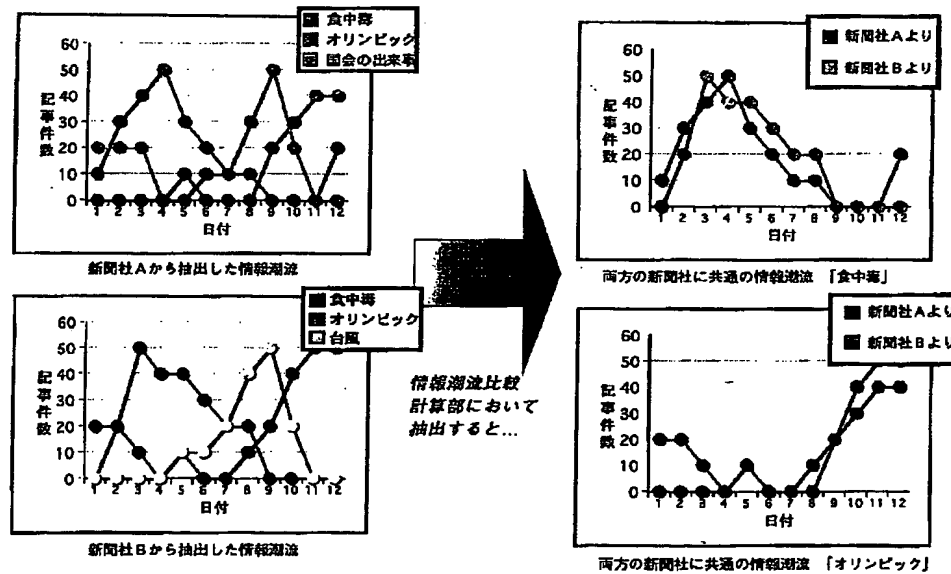


【図3】

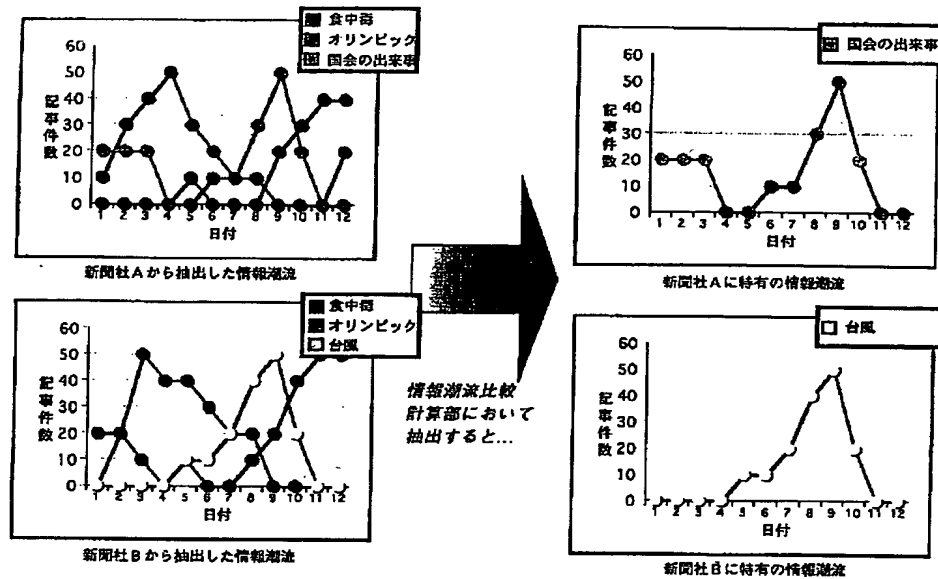


一つの情報発信媒体1日毎の文書数の変化を示したグラフ例

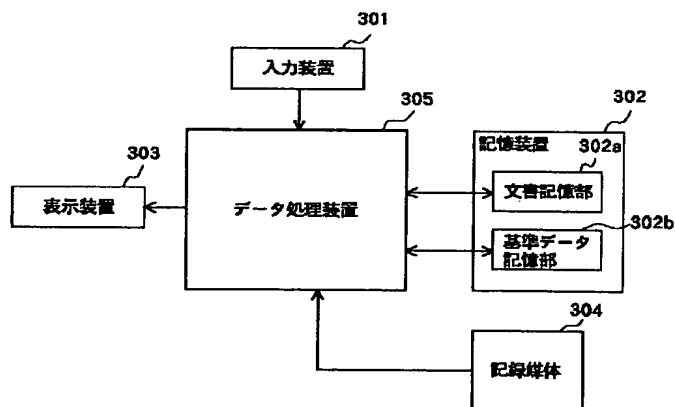
【図4】



【図5】



【図6】



フロントページの続き

(72)発明者 早川 和宏
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内